



Équivalence asymptotique des vraisemblances observée et complète dans le modèle de blocs latents

Vincent Brault, Christine Keribin, Mahendra Mariadassou

► To cite this version:

Vincent Brault, Christine Keribin, Mahendra Mariadassou. Équivalence asymptotique des vraisemblances observée et complète dans le modèle de blocs latents. XXIV èmes Rencontres de la Société Francophone de Classification, Société Francophone de Classification, Jun 2017, Lyon, France. hal-01510994

HAL Id: hal-01510994

<https://hal.science/hal-01510994>

Submitted on 21 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial| 4.0 International License

Équivalence asymptotique des vraisemblances observée et complète dans le modèle de blocs latents

Vincent Brault*, Christine Keribin**, Mahendra Mariadassou***

*Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France
CNRS, LJK, F-38000 Grenoble, France
vincent.brault@univ-grenoble-alpes.fr,
<http://www-ljk.imag.fr/membres/Vincent.Brault/>

**Laboratoire de Mathématiques d'Orsay, CNRS, et INRIA Saclay Île de France,
Université Paris-Sud, Université Paris-Saclay, F-91405 Orsay, France.
christine.keribin@math.u-psud.fr
<https://www.math.u-psud.fr/~keribin/>

***MaIAGE, INRA, Université Paris-Saclay, 78352 Jouy-en-Josas, France
mahendra.mariadassou@inra.fr

Résumé. Le modèle de blocs latents est une méthode non supervisée de classification simultanée des n lignes et d colonnes d'une matrice basée sur un modèle probabiliste de mélange. Pour estimer les paramètres de ce modèle, de nombreux algorithmes sont proposés donnant de bons résultats empiriquement mais les résultats théoriques les confirmant restent encore rares. Dans cet exposé, nous montrons que la structure particulière de ce modèle implique l'équivalence asymptotique du rapport de vraisemblance observée avec celui de la vraisemblance complète sous certaines conditions de bornes sur les paramètres et pour un régime asymptotique tel que $\log(d)/n$ et $\log(n)/d$ tendent vers 0 avec n et d . Ce résultat permet en particulier de démontrer la normalité asymptotique de l'estimateur du maximum de vraisemblance.

1 Introduction

Le modèle de blocs latents (ou LBM pour *Latent Block Model*) est une méthode non supervisée de classification simultanée des lignes et des colonnes d'une matrice de données dont le but est de mettre en évidence des blocs contrastés à partir d'un modèle probabiliste. Le LBM fait partie des méthodes de *coclustering* dont l'intérêt croît ces derniers temps grâce à leurs applications dans des domaines comme le marketing, la génétique ou encore les systèmes de recommandation.

Pour estimer les classes, il existe un certain nombre d'algorithmes fréquentistes (Govaert et Nadif, 2008; Keribin et al., 2015; Brault et Channarond, 2016) et bayésiens (Wyse et Friel, 2012; Wyse et al., 2014; Keribin et al., 2015) donnant de bons résultats empiriques. Toutefois, en dehors de Brault et Channarond (2016) qui ont montré la consistance des estimations de l'algorithme *Largest Gaps*, il n'existe pas de garanties théoriques sur la qualité des autres algorithmes.

Équivalence asymptotique des vraisemblances observée et complète dans le LBM

Des résultats partiels existent toutefois : Mariadassou et Matias (2015) ont montré la convergence de la loi conditionnelle des variables latentes pour des paramètres appartenant à un voisinage des vrais paramètres. De plus, dans le cadre du *Stochastic Block Model*, cas particulier du LBM où les matrices sont symétriques et les lignes et colonnes représentent les mêmes individus, Bickel et al. (2013) ont montré la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance dans le cas de données binaires. Pour ce faire, ils ont montré que la vraisemblance complète et la vraisemblance des observations ont des comportements asymptotiques similaires.

Dans ce travail, nous étendons ces résultats au cas de la double asymptotique du LBM pour des observations appartenant à la famille exponentielle avec un seul paramètre inconnu.

2 Modèle

Nous observons une matrice $X = (x_{ij})$ de n lignes et d colonnes et supposons que les lignes sont partitionnées en g classes de lignes et les colonnes en m classes de colonnes. Le modèle LBM suppose une structure en blocs de données obtenus par le produit cartésien d'une partition des lignes par une partition des colonnes. Plus précisément,

- les labels des lignes \mathbf{z}_i , $i = 1, \dots, n$, sont indépendants des labels des colonnes \mathbf{w}_j , $j = 1, \dots, d$: $p(\mathbf{z}, \mathbf{w}) = p(\mathbf{z})p(\mathbf{w})$;
- les labels des lignes sont i.i.d. : $\mathbf{z}_i \sim \mathcal{M}(1, \boldsymbol{\pi} = (\pi_1, \dots, \pi_g))$; de même, les labels des colonnes sont i.i.d. : $\mathbf{w}_j \sim \mathcal{M}(1, \boldsymbol{\rho} = (\rho_1, \dots, \rho_m))$. $\boldsymbol{\pi}$ (resp. $\boldsymbol{\rho}$) représente les poids du mélange pour les lignes (resp. colonnes) ;
- conditionnellement aux blocs d'appartenance $(\mathbf{z}_1, \dots, \mathbf{z}_n) \times (\mathbf{w}_1, \dots, \mathbf{w}_d)$, les observations X_{ij} sont indépendantes, de lois appartenant à la même famille paramétrique de densité φ , de paramètre ne dépendant que du bloc considéré :

$$X_{ij} | (z_{ik} w_{j\ell} = 1) \sim \varphi(\cdot, \alpha_{k\ell}).$$

Ainsi, le paramètre est $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\rho}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}$ et la log-vraisemblance des observations s'écrit

$$\mathcal{L}(\boldsymbol{\theta}) = \log p(X; \boldsymbol{\theta}) = \log \sum_{(z, w) \in \mathcal{Z} \times \mathcal{W}} \left(\prod_{i,k} \pi_k^{z_{ik}} \prod_{j,\ell} \rho_\ell^{w_{j\ell}} \prod_{i,j,k,\ell} \varphi(x_{ij}; \alpha_{k\ell})^{z_{ik} w_{j\ell}} \right)$$

où $\mathcal{Z} \times \mathcal{W}$ est l'ensemble de toutes les affectations possibles de \mathbf{z} et \mathbf{w} .

Nous supposons que la densité φ appartient à une famille exponentielle régulière mise sous forme canonique et de paramètre unidimensionnel α :

$$\varphi(x, \alpha) = b(x) \exp(\alpha x - \psi(\alpha)),$$

où α appartient à l'espace \mathcal{A} , de telle sorte que $\varphi(\cdot, \alpha)$ est bien définie pour tout $\alpha \in \mathcal{A}$. Les propriétés classiques des familles exponentielles assurent que ψ est convexe, infiniment différentiable sur $\mathring{\mathcal{A}}$, que $(\psi')^{-1}$ est bien défini sur $\psi'(\mathring{\mathcal{A}})$ et que $\mathbb{E}[X_\alpha] = \psi'(\alpha)$ et $\mathbb{V}[X_\alpha] = \psi''(\alpha)$ quand $X_\alpha \sim \varphi(\cdot, \alpha)$.

Nous posons également les hypothèses suivantes sur l'espace des paramètres :

- H_1 : Il existe une constante positive c et un compact C_α tels que $C_\alpha \subset \mathring{\mathcal{A}}$ et $\Theta \subset [c, 1 - c]^g \times [c, 1 - c]^m \times C_\alpha^{g \times m}$.
- H_2 : La fonction $\alpha \mapsto \varphi(\cdot, \alpha)$ est injective.
- H_3 : Le vrai paramètre $\theta^* = (\pi^*, \rho^*, \alpha^*)$ est dans l'intérieur de Θ .
- H_4 : α^* ne contient ni deux lignes identiques, ni deux colonnes identiques.

Ces hypothèses sont classiques. L'hypothèse H_1 permet de garantir que les poids du mélange sont raisonnablement loin de 0 et 1 pour assurer l'existence de chaque groupe. Elle permet également d'assurer que α n'est pas à la frontière de \mathcal{A} et que ψ' est lipschitzienne sur un voisinage de C_α . Les hypothèses H_2 à H_4 sont nécessaires à l'identifiabilité du modèle. Comme α est restreint à un sous-ensemble borné de \mathcal{A} , la variance de X_α est bornée.

Les hypothèses posées sont en particulier vérifiées dans le cas des variables de Bernoulli, de Poisson d'espérance bornée, et normales de variance connue et d'espérance bornée.

3 Résultat principal

Avant d'énoncer le résultat, nous définissons l'équivalence de deux paramètres.

Définition 3.1 (Équivalence) Deux paramètres θ et θ' sont équivalents s'ils sont égaux à permutations près, c'est à dire s'il existe deux permutations s et t telles que $(\alpha^{s,t}, \mathbf{z}^s, \mathbf{w}^t) = (\alpha', \mathbf{z}', \mathbf{w}')$. On note alors $\theta \sim \theta'$.

Définition 3.2 (Symétrie) θ présente une symétrie s'il existe un couple de permutations $(s, t) \neq (Id, Id)$ telle que $(\pi^s, \rho^t, \alpha^{s,t}) = (\pi, \rho, \alpha)$. L'ensemble de couples (s, t) pour lesquels θ présente une symétrie est noté $\text{Sym}(\theta)$ et son cardinal est noté $\#\text{Sym}(\theta)$.

Sans symétrie de θ , $(\mathbf{z}^s, \mathbf{w}^t)$ et (\mathbf{z}, \mathbf{w}) ont la même vraisemblance sous des paramètres θ et $\theta^{s,t}$ différents, tandis qu'avec une symétrie pour (s, t) , $(\mathbf{z}^s, \mathbf{w}^t)$ et (\mathbf{z}, \mathbf{w}) ont la même vraisemblance sous le même paramètre θ ($= \theta^{s,t}$). En pratique, l'ensemble des paramètres ayant des symétries est de mesure nulle.

Théorème 3.3 (Équivalence asymptotique) Sous les hypothèses H_1 à H_4 et pour un régime tel que $\log(d)/n$ et $\log(n)/d$ tendent vers 0 avec n et d , les rapports de vraisemblance observée et de vraisemblance complète sont asymptotiquement équivalents :

$$\frac{p(\mathbf{x}; \theta)}{p(\mathbf{x}; \theta^*)} = \frac{\#\text{Sym}(\theta)}{\#\text{Sym}(\theta^*)} \max_{\theta' \sim \theta} \frac{p(\mathbf{x}, \mathbf{z}^*, \mathbf{w}^*; \theta')}{p(\mathbf{x}, \mathbf{z}^*, \mathbf{w}^*; \theta^*)} (1 + o_P(1)) + o_P(1)$$

où la convergence o_P est uniforme pour tout $\theta \in \Theta$.

La preuve repose sur l'étude du rapport de vraisemblance conditionnelle F_{nd} et du rapport de vraisemblance profilé $\tilde{\Lambda}(\mathbf{z}, \mathbf{w})$

$$F_{nd}(\theta, \mathbf{z}, \mathbf{w}) = \log \frac{p(\mathbf{x}|\mathbf{z}, \mathbf{w}; \theta)}{p(\mathbf{x}|\mathbf{z}^*, \mathbf{w}^*; \theta^*)}; \quad \tilde{\Lambda}(\mathbf{z}, \mathbf{w}) = \max_{\theta} F_{nd}(\theta, \mathbf{z}, \mathbf{w}),$$

et distingue les cas (peu probables) de labels (\mathbf{z}, \mathbf{w}) très éloignés de la vraie configuration $(\mathbf{z}^*, \mathbf{w}^*)$ de ceux qui en sont proches. Des inégalités de concentration sont établies en appliquant une inégalité de Bernstein à des variables non bornées de variance bornée (Massart, 2007).

4 Conclusions et perspectives

Ce théorème permet de mettre en évidence les comportements similaires de la vraisemblance et de la vraisemblance complète. Ce résultat atypique et propre au modèle des blocs latents est une conséquence de la double asymptotique du modèle et peut être obtenu sous une condition faible de disproportionnalité entre les lignes et les colonnes.

À l'aide de ce théorème, nous pouvons notamment montrer la normalité asymptotique de l'estimateur du maximum de vraisemblance et la consistance de l'estimateur variationnel.

Références

- Bickel, P., D. Choi, X. Chang, H. Zhang, et al. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics* 41(4), 1922–1943.
- Brault, V. et A. Channarond (2016). Fast and consistent algorithm for the latent block model. arXiv preprint arXiv :1610.09005.
- Govaert, G. et M. Nadif (2008). Block clustering with bernoulli mixture models : Comparison of different approaches. *Computational Statistics & Data Analysis* 52(6), 3233–3245.
- Keribin, C., V. Brault, G. Celeux, et G. Govaert (2015). Estimation and selection for the latent block model on categorical data. *Statistics and Computing* 25(6), 1201–1216.
- Mariadassou, M. et C. Matias (2015). Convergence of the groups posterior distribution in latent or stochastic block models. *Bernoulli* 21(1), 537–573.
- Massart, P. (2007). *Concentration inequalities and model selection*, Volume 6. Springer.
- Wyse, J. et N. Friel (2012). Block clustering with collapsed latent block models. *Statistics and Computing* 22(2), 415–428.
- Wyse, J., N. Friel, et P. Latouche (2014). Inferring structure in bipartite networks using the latent block model and exact icl. arXiv preprint arXiv :1404.2911.

Summary

Latent Block Model (LBM) is a model-based method to cluster simultaneously the d columns and n rows of a data matrix. Parameter estimation in LBM is a difficult and multifaceted problem. Although various estimation strategies have been proposed and are now well understood empirically, theoretical guarantees about their asymptotic behaviour is rather sparse. We show here that under some mild conditions on the parameter space, and in an asymptotic regime where $\log(d)/n$ and $\log(n)/d$ go to 0 when n and d go to $+\infty$, (1) the maximum-likelihood estimate of the complete model (with known labels) is consistent and (2) the log-likelihood ratios are equivalent under the complete and observed (with unknown labels) models. This equivalence allows us to transfer the asymptotic consistency to the maximum likelihood estimate under the observed model.